

9. Using Wikipedia to Explore Issues of Systemic Bias and Symbolic Annihilation in Information Sources. Caroline Ball

Caroline (c.ball@derby.ac.uk) is Academic Librarian at the University of Derby.

Representation matters

Wikipedia brands itself as the ‘world’s encyclopaedia’ and most people accept it as such. As most educators recognise, it is usually the first place students turn for information. Yet, few people stop to think about how comprehensive it really is, how representative, how much information it fails to provide or what the real world consequences of that information gap can be.

“You cannot be what you cannot see.”

This quote is seen frequently, across a wide variety of news articles—on topics from bisexuality (Donaldson, 2019) to Indian cooking (Brehaut, 2019), baseball (Rhoden, 2018) to women in tech industries (Richardson, 2019). It does not need explanation; we immediately understand the point: representation matters. We all need role models, inspiration, and inclusion: representation. We need to be able to see examples of what we can be, what we can aspire to, the goals and opportunities open to us.

Just as you cannot be what you cannot see, you cannot learn what you cannot find. Or perhaps you *can* learn something, but not what you were looking for. You can learn that you are unimportant, that you do not matter, that your experience, your language, your culture, your existence is unimportant. The existence of an article on Wikipedia gives that subject weight. So what happens when there is no Wikipedia article on a person or place?

The quote at the beginning of this section is rarely accurately attributed to its original source: Marian Wright Edelman, an American activist for children’s rights. You can look her up on Wikipedia. Her article is brief, a mere 1,438 words, including references. By contrast, the article on professional footballer Cristiano Ronaldo is 33,108 words, including references. This in itself tells us something about relative importance in a world of crowd-sourced information!

As a direct result of its crowd-sourced nature, Wikipedia is subject to the systemic bias that exists in our societies, and in turn contributes to

symbolic annihilation. However, Wikipedia's failings in this regard can be turned into a benefit by using it as a tool to raise awareness of these issues in the classroom and involve students in actively working to improve coverage.

What is systemic bias?

Systemic bias refers to the everyday practices and processes embedded within systems or institutions that can create or support disadvantageous outcomes for certain groups and/or individuals from those groups. Quite frequently, these processes will have been designed by the dominant group/s in a given society or culture, assuming that their experience is the norm or default, without any awareness (or consideration) that other individuals do not experience the world in the same way. These practices and processes can then, to anyone outside of those dominant groups, become barriers to access, participation, or, at the most extreme level, to existence itself.

Caroline Criado-Perez's recent book 'Invisible Women' (Criado-Perez, 2019) is an excellent source of examples of systemic bias as it relates to gender-related data, via the 'gender data gap'. She discusses how failing to gender-differentiate data (by using a universal default that is often 'average male' in size and shape, or by not involving women in data collection or research) can unintentionally create difficult, harmful, or even fatal outcomes for women. For example, women as a gender irrespective of race or ethnicity are more likely to die from heart attacks, because their symptoms are not 'typical' - or more accurately, are not the same as symptoms experienced by men. Most women cannot use their smartphones single-handed because phone size was designed to fit the average male hand span. Women are more likely to be seriously injured or die in car accidents, because the safety tests use crash test dummies based on the average male body size and shape, which is significantly taller and heavier than the average woman. As a further failing, much data is not differentiated by race or ethnicity either, which can serve to even further reinforce systemic bias.

These are extreme examples of systemic bias in action, but in many ways systemic bias in information sources can be just as damaging in ways that may not be as immediate and visible. In the case of Wikipedia, because of its crowd-sourced nature, individual biases, conscious or otherwise, can become systemic biases. These biases then play a huge role in shaping the

creation of Wikipedia articles overall, leading to major inconsistencies in coverage and quality, as we will see later in this chapter.

What is symbolic annihilation?

Symbolic annihilation is a concept first articulated almost fifty years ago by George Gerbner and Larry Gross (Gerbner & Gross, 1976), describing the complete absence or minimal representation in the media of certain groups of people (frequently based on race, gender, sexual orientation, and/or socio-economic status). This lack of visible representation serves to maintain social inequality by excluding those groups that are deemed socially insignificant from the social and cultural narrative.

The concept was originally interpreted specifically in terms of absence, but in the intervening years the concept has been further refined to now accommodate three elements, all of which can be equally damaging to individuals and groups: omission, trivialisation and condemnation (Tuchman, 1978). Some forms of representation can be worse than absence if they consist solely of offensive or inaccurate stereotypes, or negative or comical portrayals.

How might this ‘symbolic annihilation’ affect an individual’s understanding of their place in society, history and culture? Might this absence potentially lead to feelings of alienation and lack of worth? What happens “when someone with the authority of a teacher describes our society and you’re not in it?” (Rich, 1993, p. 16).

Various studies on the issue and impact of symbolic annihilation have been undertaken in the years since Gross and Gerbner first identified the concept—surveying Native Americans to assess the impact of their “actual and symbolic annihilation” (Merskin, 1998, p. 335); analysing plantation museums for their depiction of the lives of enslaved African-Americans (Eichstedt & Small, 2002); assessing animated cartoons’ portrayals of marginalised groups such as LGBTQ+, women, the elderly and racial minorities (Klein & Shiffman, 2009); even evaluating representations of LGBTQ+ individuals in *Star Trek* (Venzo, 2016), to name just a few.

Common themes emerge from many of these research studies, articulated by an unnamed Apache male in Merskin’s study of Native Americans:

“Most people believe the generalizations. This certainly creates an identity crisis for many.... [They] often find themselves not knowing who to identify with. How realistic these portrayals are goes beyond historical

accuracies and reach into individual beliefs about the self’ (Merskin, 1998, p. 339).

In 1976 Gerbner and Gross were solely applying the concept to visual media such as television, and many of these subsequent studies have also focused on visual media portrayals. However, there is a strong case to be made for extending the concept from visual media portrayals into textual and informational contexts. Recent studies focusing on the role of community archives in combating symbolic annihilation and enhancing representation have focused on the ‘epistemological impact’ of such endeavours in changing “the nature of *what* can be known about a community’s history and *how* it can be known” (Caswell, Migoni, Geraci, & Cifor, 2017, p. 17). As one of the world’s largest information sources, there is surely relevance for Wikipedia here.

How do these issues manifest in Wikipedia?

It is common to see news and media articles declaring ‘Wikipedia has “banned” this or that, but this type of claim represents a fundamental misconception of how Wikipedia works—that is, the misconception that Wikipedia as an organisation that acts with a single will and voice. Wikipedia is made up of millions of individual editors, who all bring their own attitudes, opinions, biases, prejudices, and beliefs to the task of creating and editing articles.

Wikipedia describes the average editors of the English-language Wikipedia as young, white, college-educated males, technically-inclined, living in majority-Christian, Northern Hemisphere countries (“Systemic bias”, n.d.)—a group that has been described in the media as “a bunch of male geeks who are wealthy enough to afford a \$2,000 laptop and a broadband connection” (Montellaro, 2015). Recent surveys have estimated that only 8.5%–16% of Wikipedia editors are female—even fewer are people of colour of any gender (Smith, 2015). Indeed, it is striking how closely Gerbner and Gross’ original description of the characters that dominated the screen in 1970s television in their original description of the concept of symbolic annihilation fits the majority of today’s Wikipedia editors: “three quarters [...] are male, American, middle- and upper-class, and in the prime of life” (Gerbner & Gross, 1976, p. 183).

Wikipedia was not ‘designed’ to operate the way that it does; in reality, as a crowd-sourced encyclopaedia, it was not designed at all. Its processes and procedures have evolved exclusively due to the input of those who

contribute to it. However, if those individuals do not represent the broad spread of humanity (and as already seen, they clearly do not), Wikipedia can begin to reflect those individuals' conscious or unconscious biases, which then become fixed in place as systemic bias.

One of the major challenges involved with broadening the base of Wikipedia editors is that it is not only contingent on inclination or ability, but also on access—both to the Internet itself and to sources of information—and this access varies according to geographical location (not to mention socio-economic status). A study undertaken in 2011 (Graham, Hale, & Stephens, 2011) compared the percentage of 'geotagged' English-language Wikipedia articles to world populations broken down by geographical region. It found an overwhelming bias towards Europe and North America, with those two regions accounting for 84% of English-language Wikipedia articles (Graham et al. 2011). It is no coincidence that these regions also have the highest Internet penetration rates in the world, with North American averaging 77% and Europe between 79% (Euro area) and 81% (European Union) (The World Bank Group, n.d.).

With such a non-representative selection of the world population responsible for the vast majority of creation of and edits to articles in the English-language Wikipedia, it is no surprise that the content itself is also not representative. Articles about notable women are under-represented (Leonard, 2018). Coverage relating to Africa, Latin America, and the Middle East in the English-language Wikipedia is rated by Wikipedia itself as poor to mediocre ("WikiProject: Countering systemic bias", n.d.), and those articles that do exist are often written from a European or North American perspective. Articles on 'universal' topics often fail to include examples from these regions as well—do people in African countries not eat lunch, for example (Lunch, n.d.)?

We can see even from these brief examples the elements of symbolic annihilation at play: *omission* (articles on important individuals, regions, cultures, and topics can be missing altogether), *trivialisation* (articles can vary greatly in length, coverage, and quality) and *condemnation* (articles can frequently represent a Western cultural viewpoint, often to the detriment of other cultures).

This narrow, homogenous editor pool situated within specific geographic regions raises further issues beyond motivation and access, all of which serve to further embed systemic bias within Wikipedia's processes and guidelines for editors. Such issues include the nature, focus, scope, and

format of the material available to editors on which to base the articles they write and edit.

Access to sources of information becomes an issue—and not just any sources of information, but those sources that Wikipedia deems more ‘reliable’, usually meaning traditionally published material such as newspapers, books, and academic journals. Original research is not permitted on Wikipedia, an understandable policy that aims to restrict fringe theories and unverifiable claims, but it also means that content for articles is restricted to what is deemed ‘publishable’, a concept subject to a whole range of external forces, few of which are devoted to ensuring equity and visibility for marginalized groups.

Another factor is the Anglo-American domination of the educational and academic publishing industry. The United States and the United Kingdom come second and third in terms of the number of books published per year, behind China (Ingenta, 2014). In terms of book sales, the United States accounts for 26% of world book sales, with the European Union making up another 33%. When focusing exclusively on digital sales, the Anglo-American domination is even more striking—12.5% in the UK, 15% in Canada and 20% in the United States (Centeno, Lara, & Vallejo, 2014). Despite the fact that English comes a distant third in the number of native speakers, behind Spanish and Mandarin Chinese (Ethnologue, n.d.), non-English journals are frequently excluded from the high-status journal indexes, increasing pressure on non-native English speakers to publish in English in order for their research to be widely-known and recognized (Curry & Lillis, 2018).

Wikipedia guidelines also do not consider oral knowledge a reliable source, which can greatly reduce the inclusion of much material relating to Indigenous cultures, which have traditionally valued oral transmission of cultural knowledge over written transmission. Concerns generally focus on the difficulty in checking oral citations for accuracy and the lack of academic authority involved, although both of these issues, and others, have been criticised as displays of ‘cultural imperialism’, valuing one (dominant) culture’s means of knowledge transmission over another’s (Gallert & van der Velden, 2015).

The selection of topics that subjects editors choose to write about is another area fraught with issues of perspective and bias. Wikipedia does have guidelines relating to the issue of ‘notability’ (i.e. whether a topic is considered important enough to justify a stand-alone Wikipedia article).

However, who determines whether something is notable enough to be 'Wiki-worthy'? The Wikipedia editor community does, but as we have already seen, this community is not global nor representative in a myriad of ways. What is deemed worthy by a young, white, male, educated American or Northern European may bear no relation to those issues deemed important by a middle-aged woman from Grenada, a teenage girl from Estonia, or a grandfather from Chile.

The issue of notability is one frequently used against female subjects on Wikipedia. Detailed articles about women are often rejected for not being considered 'notable' enough, yet there are a great number of very short articles about men. Many critics argue that female subjects on Wikipedia are being held to a higher notability standard than male subjects, bringing to mind the famous quote from Charlotte Whitton (she has a Wikipedia page, look her up!): "Whatever women do they must do twice as well as men to be thought half as good" (cited in Powell, 2018).

The problem of systemic bias goes beyond Wikipedia. It is inherent in the very systems we use to inform ourselves—including Wikipedia, of course, but also the sources Wikipedia relies upon, the Internet systems we use to access those sources, the languages that material is published in, and the educational, cultural, and political trends that influence and control what is published and what research is undertaken.

If a subject—whether it be a person, a concept, an artefact, or a place—is not valued, it is not researched nor written about. If no one is writing about it, or is writing in a language other than English, Spanish, or Chinese, it gets little visibility. If something is not published widely enough, the information it contains cannot be disseminated, digested, synthesised, and reproduced for a Wikipedia article. If it cannot be found or discovered, it cannot be referenced in a Wikipedia article. If information about a subject in a Wikipedia article does not cite verifiable material, it is marked for deletion.

It is a vicious cycle that brings us back to Cristiano Ronaldo and Marian Wright Edelman. One is a professional footballer, the other is an activist for children's rights. In theory, we may know which role we would like to think our societies value the most, and yet, their respective Wikipedia articles do not reflect that relative value.

Teaching with Wikipedia

My own experience with Wikipedia as a teaching and learning tool began with an editathon for International Women's Day, aimed at improving Wikipedia's coverage of female biographies. I had dabbled with editing Wikipedia myself, teaching myself the basics, but this was the first time I had used it in an educational context. Students attending the webinar responded with enthusiasm, and as I was supporting them in writing and editing articles and doing research to find missing citations, it dawned on me how useful a tool this could be in my own line of work.

As an academic librarian working in a university, my primary focus is on enhancing students' information and digital literacy skills, and Wikipedia seemed an ideal tool for this. Many of the elements of digital literacy I teach and support are required when editing and writing Wikipedia articles: writing and copy-editing, research and referencing, source evaluation, critical thinking. I began to incorporate Wikipedia into my library workshops: using it as an example when teaching referencing, demonstrating the 'citation needed' tags in articles; showing students the reference lists at the end of articles; explaining how it functioned, how articles were organised into quality categories, when to use it and when to skip it.

This eventually led to the opportunity to create an entire digital capabilities module structured around the use of Wikipedia. The module included two assignments—an individual assignment and a group assignment—requiring students to create portfolios of articles edited and created from scratch. Through the course of the module, students had classes on referencing, copyright and plagiarism, research, source evaluation, media literacy and peer review, all linking back to Wikipedia.

Such an extended, in-depth examination of Wikipedia afforded the opportunity to critically examine its weaknesses as well as its strengths. As the module progressed and the students' understanding of Wikipedia became deeper and more nuanced, its flaws became more visible to both students and teacher. Students started to notice gaps in the coverage, commenting on their surprise and disappointment when articles were not as comprehensive as they'd hoped or failed to cover issues they considered important. They also expressed frustration at their inability to find reliable, accurate information to improve articles on less high-profile or mainstream topics.

I felt that it was important to address these issues head-on with the students, by making them more aware of the context in which Wikipedia operates. Some of the activities described below were therefore designed specifically for inclusion in the digital capabilities module, while others were later designed for inclusion within more general library workshops or to be standalone activities. For example, the quality sampling activity was used within a library workshop focusing on source evaluation—we discussed how articles within Wikipedia rely on reliable, high-quality sources in order to achieve higher quality ratings, what happens if these sources are not available, and why that might be. An editathon, on the other hand, works as a stand-alone activity in itself.

The activities detailed below are all activities I have used with students in both small and large groups. They can be used by librarians or teachers, on their own or as part of a wider sessions. The lengths of time required are a rough estimate, as timings can depend on a variety of factors, including class size, and the layout and facilities of the room available.

Activities

Representation Hunt

Learning outcomes:

- Understand concepts of systemic bias and symbolic annihilation in information sources
- Explore issues of diversity and representation in print-based media
- Consider how information sources can ignore or misrepresent individuals and/or groups

Materials:

Representation statements

Print-based media (magazines, catalogues, newspapers, books)

Paperclips

Time:

30-45 minutes

Description:

A ‘representation hunt’ can be a useful print-based activity to introduce students to the concepts of systemic bias and symbolic annihilation in information sources, before moving on to apply those same concepts to the digital world of Wikipedia.

This is a relatively simple, but highly effective exercise, although it does require an element of preparation beforehand in gathering materials. A large sample of disposable print-based media is required—these can be catalogues, magazines, newspapers, books—on any topic at all. A wide variety is ideal, but whatever is available can be used. The amount required will inevitably vary on the size of the class involved.

A number of ‘representation statements’ need to be prepared, with some form of identifying statement relating to gender, ethnicity, job, disability, hobby, or interest. Examples can include: ‘I am a girl in a wheelchair’, ‘I am a trans woman’, ‘I have parents of the same gender’, ‘I am a female firefighter’, ‘I am an overweight man’, ‘I am a boy who likes pink’, ‘I am a slim woman’, ‘I am Native American’, etc. These can be printed on sheets and cut into individual slips to hand out to students.

Students are each given a number of statements to look for. The number can depend on the amount of materials available and the number of students: 4-6 is good to start with, either per student or in pairs or small groups, depending on class size. The students are given 5-10 minutes to work through the print materials available to them looking for references to individuals who match the representation statements, whether visual (images) or textual (references within text). When they find a match, they should tear out the page and clip or staple the representation statement to it.

Students can get frustrated during this activity if they fail to find any matches for their representation statements. However, this in itself can provide useful material for subsequent discussions about how it must feel to be personally on the receiving end of that lack of representation. Students can be divided into small groups (if they are not already, and again, group size depending on overall class size) to discuss which individuals seem more represented than others and what messages are being sent to those not represented at all as a result.

The next stage of the activity is to look at the context in which the representation occurs, taking one or two examples identified in the previous activity and discussing them in small groups (these groups can be

the same or different in composition to the previous discussion activity). This is a useful stage to introduce students to the three elements of symbolic annihilation identified earlier in this chapter: *omission*, *trivialisation*, and *condemnation*. For example, are the individuals matching the representation statements the primary focus of the text or image, or secondary/background to another individual? Is the context positive or negative? Do the individuals matching the representation statements have agency or are they dependent on the direction or assistance of another? Does the representation rely on offensive or inaccurate stereotypes?

The aim of the session is to demonstrate practically just how lacking in diversity most mainstream information sources can be, and how difficult (if not impossible) it can be for some individuals to find themselves represented accurately and positively in the media. This can be a useful springboard to get students thinking about these issues in terms of print-based media, before applying the same issues of representation and accuracy (and lack thereof) to Wikipedia.

Wikhopping

Learning outcomes:

- Explore gaps in coverage across Wikipedia articles
- Assess variations in content across Wikipedia articles

Materials:

Computers

Pen and paper

Time:

15 minutes

Description:

Wikipedia has a 'Random article' link in its menu bar, which generates a random article from the 5,957,364 articles (as of writing) in the English-language Wikipedia. Students are given an allotted length of time to browse (keep this relatively short, no more than 5-10 minutes) and asked to keep a log of the articles randomly generated by this tool (see below for examples).

Criteria identified for monitoring can vary depending on what elements, omissions, or evidence of bias are under scrutiny in the class. Examples could include assessing gender bias by monitoring the balance of male versus female subjects, and the comparative length and quality of articles—or the even more striking lack of visibility for non-binary and transgender subjects.

Alternatively, students could look for evidence of racial or geographical bias by assessing whether certain countries or regions are more represented than others, or whether ‘universal articles’ contain omissions or sparse information relating to those regions or cultures.

| Article title | Word count | Article rating | Male v Female | Location | Notes |
|-------------------------|-------------------|-----------------------|----------------------|-------------------|----------------------------|
| Quarter (Canadian coin) | 1210 | C | n/a | Canada | |
| Ernest Kombo | 165 | Stub | Male | Republic of Congo | |
| Lindy's Sports | 83 | Stub | n/a | USA | Male publisher/ founder |
| Anjalin | 58 | Stub | n/a | Iran | |
| Meteorite Island | 198 | Start | n/a | Greenland | |
| Tāwhirimātea | 748 | Start | n/a | New Zealand | |
| Hamid Ghodse | 1319 | Start | Male | Iran | |

| | | | | | |
|-----------------------------|------|-------|------|--------|--|
| Max Corput | 659 | Start | Male | | |
| Rufoclanis rosea | 166 | Stub | n/a | Africa | |
| The Decoy (1935 film) | 170 | Stub | n/a | France | |
| Fujicolor Pro | 816 | Start | n/a | Japan | |
| Long Island Creek | 112 | Stub | n/a | USA | |
| Arrow Lakes | 1215 | Start | n/a | Canada | |
| Sugar Creek Slavic Festival | 1242 | Start | n/a | USA | |

Fig 1. Table showing an example log from a ‘Wikhopping exercise’, detailing the random articles generated, their length, quality, gender and geographical location.

This can be a useful and quick means of introducing students to the wide variations in coverage within Wikipedia, the elastic definition of ‘notability’ (witness how short some of the male biographies are), and the lack of coverage in some areas. For instance, in a random sampling of fifteen articles, only one article related to the entire continent of Africa, and there were more articles on bodies of water in North America than there were on women!

An alternative approach focusing exclusively on gender disparities would be to see how many clicks of the random article generator it takes to arrive at a biography about a non-binary or transgender subject. As the table below demonstrates, frequently such an exercise results in no such articles at all.

| # of clicks | Subject | Country | Gender | Word count |
|-------------|------------------------|-------------|--------|------------|
| 1 | Jacques Thuillier | France | Male | 419 |
| 5 | Margrit Thommen | Switzerland | Female | 54 |
| 7 | Navin Bhakta | India | Male | 93 |
| 8 | Honorio Pueyrredón | Argentina | Male | 229 |
| 9 | Borja Criado | Spain | Male | 512 |
| 10 | Mustapha Skandrani | Algeria | Male | 526 |
| 11 | Richard Venture | American | Male | 298 |
| 15 | Robert Jarvik | American | Male | 925 |
| 17 | Peter Badham | English | Male | 283 |
| 20 | Raad Mutar Saleh | Iraq | Male | 104 |
| 26 | Farid Zhanigirov | Russia | Male | 28 |
| 30 | Donald Kenney | American | Male | 211 |
| 32 | Kristina Paner | Philippines | Female | 4069 |
| 33 | Carlo Raimondi | Italy | Male | 293 |
| 36 | Nicola Correia-Damude | Canada | Female | 537 |
| 42 | Royce Hunt | Australia | Male | 442 |
| 43 | Iván Varga | Argentina | Male | 276 |
| 44 | Javier Araújo | Colombia | Male | 88 |
| 50 | Slobodan Mistic-Brenda | Canada | Male | 302 |

Fig 2. Table showing the results of 50 clicks of the random article generation, monitoring gender and geographical location.

From the random sampling of 50 articles seen in Fig. 2, we generated 19 biographical articles—only 3 of which were about women. Football players

had a greater representation in our sampling than the entire female gender. Africa and the Middle East were represented by 1 article apiece, whilst Europe was represented 5 times and North America 5 times.

This approach is quite simple and easy to undertake with a group of any size. A random sampling of articles can clearly demonstrate both how underrepresented certain groups and regions are within Wikipedia and how brief some of the articles on already underrepresented groups can be. This ties back into the issue of notability raised earlier in this chapter and how it can frequently be used against female subjects and subjects from certain regions of the world.

Quality sampling

Learning outcomes:

- Recognise variations in coverage across globally comparable topics
- Learn to identify potential signs of systemic bias

Materials:

Lists of topics for comparison

Computers

Pen and paper

Time:

30-45 minutes

Description:

In this activity, students are provided with lists of comparable topics from different cultures and countries around the world—for example, capital cities, rivers, heads of state. There is no set requirement for the content of these lists, although a topic that is truly global in scope and representation works best.

Students then look up the Wikipedia articles for these topics and compare the length and quality of the articles. It may be appropriate to ask the students to anticipate ahead of time, using their growing knowledge of systemic bias in information sources, which subjects might have the most

detailed articles, and which the least, and then compare their guesses against their own data table.

| Country | Capital city | Word count | Quality |
|----------------|---------------------|-------------------|----------------|
| Ankara | Turkey | 11,447 | C-class |
| Berlin | Germany | 15,666 | B-class |
| Bogota | Colombia | 14,728 | C-class |
| Buenos Aires | Argentina | 21,362 | B-class |
| Cairo | Egypt | 10,461 | B-class |
| Delhi | India | 15,755 | GA-class |
| Dhaka | Bangladesh | 9,609 | FA-class |
| Havana | Cuba | 12,425 | B-class |
| Karachi | Pakistan | 15,164 | B-class |
| Kinshasa | DR Congo | 6,316 | C-class |
| Lagos | Nigeria | 9,021 | B-class |
| Lima | Peru | 11,167 | B-class |
| London | England | 21,966 | GA-class |
| Luanda | Angola | 4,207 | Start-class |
| Madrid | Spain | 17,250 | B-class |
| Mexico City | Mexico | 19,987 | B-class |
| Moscow | Russia | 20,365 | B-class |
| Nairobi | Kenya | 10,477 | C-class |
| New York City | United States | 33,616 | B-class |
| Santiago | Chile | 12,973 | C-class |
| Sao Paolo | Brazil | 19,099 | B-class |

| | | | |
|-------------|----------|--------|----------|
| Shanghai | China | 17,371 | GA-class |
| Tegucigalpa | Honduras | 12,316 | B-class |
| Tokyo | Japan | 9,642 | B-class |
| Toronto | Canada | 16,277 | B-class |

Fig 3. Table showing list of capital cities and the relative length and quality of their articles.

In the example shown in Fig. 3, we can see there are great discrepancies in length—compare Luanda’s 4,207 words to New York City’s 33,616 words. There is also a wide variety in terms of the quality of these articles, an element that is often connected to length and detail. It is no coincidence that Luanda as the shortest article is also the lowest-rated, whilst articles on the capitals of nations such as China, India and the UK achieve one of the highest ratings, that of GA (Good Article).

An alternative approach is to look at the top-rated articles within the English-language Wikipedia (those awarded FA or Featured Article status, i.e. those articles that make it to the front page of Wikipedia) and assess how representative they are. This task is not as onerous as it may sound; the current number, as of writing, is a mere 5,672 articles, less than 1% of all Wikipedia articles. Wikipedia lists these Featured Articles by category, so students could be assigned a particular category to assess.

For example, the category for biographies in the field of art, architecture and archaeology contains 47 articles—of these, only 13 are about female subjects. In the biographical categories for ‘Business, economics and finance’, ‘Chemistry and mineralogy’, ‘Engineering and technology’, ‘Medical’, ‘Philosophy and psychology’, ‘Physics and astronomy’, there is not a single article about a female, non-binary or transgender subject.

These categories can also betray the Western cultural bent of Wikipedia editors. In the category of ‘Religion, mysticism and mythology biographies’, out of 66 articles, 62 are male, 63 are Christian, and only 3 hail from outside Europe and America (one of whom is Jesus!) There is not a single Featured Article relating to Islam, Hinduism, Buddhism, Sikhism, or Judaism. Given that collectively these religions have more adherents than Christianity, this is clearly indicative of the bias within Wikipedia towards Western cultural topics.

This approach can reinforce previously explored issues of systemic bias by making students aware that even where there is coverage, or where topics are addressed, there can be vast variations in the levels of detail and the quality of the material. This can reflect both the level of interest in the subject on the part of the, as previously discussed, unrepresentative editors, and the relative availability or scarcity of English-language sources on a subject.

Sentiment analysis

Learning outcomes:

- Recognise how editing choices can signal author's opinion or attitude
- Apply sentiment analysis techniques to Wikipedia articles

Materials:

'List of controversial issues' page (or equivalent)

Word processing software

List of positive and negative words

Time:

30-45 minutes

Description:

Sentiment analysis is the process of analysing natural language to determine the emotional character of the content, via examination of the words used and the context in which they are used. Effectively it permits analysis of an author's attitude towards something or someone, whether that be positive or negative, overt or subtle.

In the era of social media, sentiment analysis has become a hot topic, with many tools developed to analyse posts on platforms like Twitter and Facebook. The analysis is largely done via computer software, using algorithms that compare words in a given sample of text to a list of words designated as positive (e.g. honest, accomplished, peaceful, impressive) and a list of words designated as negative (e.g. corrupt, violent, angry, bad). These tools also look at the context of the identified terms for the

presence of negation that can change the polarity of a word (e.g. ‘not’ good) and often add greater weighting for variations in positivity and negativity (‘excellent’ given greater weighting than ‘good’, ‘evil’ greater than ‘bad’).

There are free sentiment analysis tools available online; however, without a detailed understanding of the algorithms involved and which words have been assigned positive or negative weighting, these tools can be unreliable. Several of these tools were tested by the author with the same piece of text (the introduction from the Wikipedia page for the British National Party) and they gave wildly varying results—from strongly negative to strongly positive. This type of example in itself is an indication of how systemic bias can manifest in software and could be used as an example of how programmed technology will inevitably reflect the world view and biases of its programmer(s).

However, a rudimentary version of sentiment analysis can be performed manually on short passages of text, such as Wikipedia articles. This activity can be a useful way to demonstrate to students that any piece of text can be subject to elements of bias, however subtle, via the way information is presented and the language used, even when the author may be striving for neutrality or objectivity. One of the key elements of systemic bias is that it is frequently unconscious and can be exhibited even by authors with the best of intentions.

This activity works best as a demonstration when using topics more likely to betray an easily recognisable level of bias or lack of neutrality, such as biographies, politics, organisations, etc. Wikipedia maintains a ‘List of controversial issues’ page, which can be a good place to start to identify articles for students to analyse. Students can either make their own choices or the instructor can choose for them; the latter approach may be best, given the tendency of individuals to veer towards topics they are knowledgeable about or can identify with.

Sentiment analysis of Wikipedia articles can take several different approaches, and these can be used individually or in combination. A table or checklist can be provided to students to guide them in their analysis (see Fig. 4 for an example), or they can be asked to use their own judgement when evaluating an article.

An initial approach can be to analyse the article as a whole from the perspective of content and structure. For example, what information is included in the ‘lead section’ part of the article above the contents box?

This is the first (and sometimes only) part of any article that many users read, so it is intended to serve as an introduction to the article and a summary of the most important information. If this summary slants more heavily towards negative or critical material, this can leave users with a negative impression of the overall topic.

Which facts or information is included in, or excluded from, an article can be indicative of the overall tone as well. For instance, in a biographical article, if there is lengthy or frequent reference to positive material such as awards, achievements, positions, ranks, or charitable works, but little mention of criticism, controversies, or scandals, this is potential evidence of the biases of the editors behind the content.

The structure of a page can also play a significant part in contributing towards a negative or positive impression. If an entire section of the article focuses on positive or negative elements and is immediately visible in a page's contents box, whilst contrasting information is not highlighted but rather 'buried' in a body of text, this too can mislead or sway readers in their impression of the topic.

| | YES | NO | COMMENTS |
|--|-----|----|----------|
| Does the lead section take a neutral tone? ('No' if clearly positive/negative material included) | | | |
| Does the structure betray any bias, i.e. entire sections devoted to positive/negative elements? | | | |
| Are some sections longer than others, despite being equally important? | | | |
| Does it represent competing viewpoints equally? | | | |
| What are the sources used? Are they impartial and reliable? | | | |
| Does it contain statements that lack verification? | | | |

| | | | |
|--|--|--|--|
| Does it contain subjective ‘value statements’ (i.e. ‘the best’, ‘the most important’?) | | | |
|--|--|--|--|

Fig 4. Checklist for students to assist in assessing a Wikipedia article for bias

Let us use the Wikipedia article for the ‘National Rifle Association’ (NRA) as an example. The lead section for this article refers to frequent and sustained criticism of the NRA from a variety of groups and specifically mentions several high-profile school shootings. The structure of the page includes an extensive ‘Criticism’ section subdivided by topic. By contrast, the ‘Programs’ section of the page, which makes reference to more positive initiatives by the NRA to promote firearm safety is much shorter and is not subdivided by topic. The ‘Public opinion and image’ section (the existence of which alone suggests there is an issue of bias to be discussed) is largely weighted towards criticism.

The Talk page for an article—an administrative page where editors discuss changes/updates made to the article—is another good place to guide evaluation of an article. It will contain Wikipedia’s own internal rating of the article and will also display any warnings regarding issues surrounding the article. For example, on the NRA Wikipedia article, the controversial warning reads “The subject of this article is controversial and content may be in dispute. When updating the article, be bold, but not reckless. Feel free to try to improve the article, but don’t take it personally if your changes are revered; instead, come here to the talk page to discuss them. Content must be written from a neutral point of view. Include citations when adding content and consider tagging or removing unsourced information” (Talk: National Rifle Association, n.d.).

Another method of assessing the perspective of an article is by using a visual coding approach, which helps students see how much negative or positive language or information is included in the article. This approach can also serve as a useful way to introduce students to the basics of coding qualitative data, a skill that may be required by some in future should they pursue dissertations or research projects.

A list of positive and negative words is useful to have to hand. A number of variations of these are available freely online, although for the purposes of this activity the author used Liu and Hu’s Opinion Lexicon (Bing & Mingqing, 2004). This is very long (around 6,800 words), so it is not recommended that students use these for constant consultation when

analysing their chosen articles. Rather, they are useful to provide as an exemplar of the kind of language they are looking to identify.

To do this visual coding, students begin by copying sections of the article into a document. They read through the text carefully and then, using a highlight tool, highlight phrases and statements as either positive (green), neutral (leave un-highlighted), or negative (red). Positive and negative text can refer both to the individual words used and the overall point or context of the text. Assessing the content of a Wikipedia article in this way—by delving into the actual text, facts, and language used—can demonstrate to students that even when an article exists, even when it may appear to be detailed and lengthy, it can still have significant issues of systemic bias and symbolic annihilation, depending on the information editors have chosen to include or exclude and emphasise.

Editathon

Learning outcomes:

- Increase knowledge of Wikipedia
- Increase information and digital literacy and critical thinking
- Develop research and writing skills
- Work to address issues of systemic bias

Materials:

Computers

Wikipedia editors

Time:

30-45 minutes

Description:

More than any other approach with Wikipedia, actively getting students involved in editing is the best way to introduce and teach them about how Wikipedia functions. An 'editathon' is an event at which Wikipedia editors come together to edit and improve specific topics, which are usually chosen in advance. They can be open to all levels of editors, from

experienced Wikipedians to complete beginners, and usually involve basic editing training. Editathons can last from a few hours to all-day events—the *Museo Soumaya* in Mexico City currently holds the Guinness World Record for the longest editathon—72 hours (‘Longest edit-a-thon | Guinness World Records’, n.d.).

Some editathons are organised as part of an established project dedicated to addressing a particular issue: Women in Red, for example, which aims to increase the number of female subjects in Wikipedia’s biographies (which has improved from 15% to 18% since the project began). Art+Feminism works to improve coverage of articles relating to gender, feminism, and the arts. The African 10,000 Challenge is aiming to reach 10,000 article improvements for Africa.

Themed editathons are one of the best ways to introduce students to the concepts of systemic bias in Wikipedia by working to actively address those imbalances. They can either aim to create new articles to improve coverage in particular areas or focus on editing or improving existing articles. Either approach is an excellent way to bring the issues discussed in this chapter to life for students, by getting them actively working to do their part to address the problem. Wikipedia provides an excellent guide on the details and practicalities of how to run an editathon (‘Wikipedia’, 2019).

It is advisable to select a particular theme for your editathon, although it doesn’t have to link in to one of the established WikiProjects. A themed approach ensures a clear goal and provides boundaries within which the students can work. With completely open editathons, the challenge is that students can be overwhelmed by the number of articles in need of improving or creating, and can find it difficult to select a topic or subject to work on. Wikipedia keeps a list of ‘missing articles’ broken down by category that can be used as a starting point from which students select topics. It is recommended, particularly with new student editors, to select a category relevant to their module, programme, or assignment.

A seasoned Wikipedia editor is essential for an editathon to provide training and guidance for new editors. If you do not have the expertise within your own institution, you can get in touch with your local Wikimedia chapter for support. Wikimedia is always looking for educational partners to work with and can put you in touch with Wikipedia volunteers in your region who may be able to help. A ratio of one Wikipedian per 10 students is ideal, although this is not always possible. A

research expert (like a librarian) is also recommended, to support students in researching their chosen topics.

Whilst editathons can last for as long or as little as need be, it is advisable to be realistic in your approach, particularly when trying one for the first time. You need to allow enough time for new editors to become familiar with the basics of editing Wikipedia and then to put those new skills into practice. An hour would not be long enough. Somewhere between 2-3 hours would be ideal for a first editathon with new editors.

If you have a group of participants with mixed experience, setting up designated areas within your space, each with a Wikipedia editor on hand, is recommended. That way one group can get started with setting up accounts and training, whilst another group with more experience can be editing with limited input required from trainers. Alternatively, you may want to try setting up areas or groups focusing on different tasks—copy-editing vs. new article creation. Whatever the approach, setting up your space so that editors work in groups is recommended—this provides an opportunity for communal help and creates more of an enjoyable, social atmosphere.

Editathons, whilst serving as an excellent tool to actively address and combat issues of systemic bias and symbolic annihilation, also deliver a range of other benefits to students. Writing and editing Wikipedia articles helps students develop their reading and writing abilities. Adding references to provide evidence for statements in articles utilises students' research skills in finding material and enhances their awareness of the importance of citing and referencing in academic work. Adding images and other digital media ensures that students are aware of copyright and related licensing schemes such as Creative Commons. Because of the community-based nature of Wikipedia editing, students' work will be peer reviewed by other editors and they will receive critical feedback on their edits, either directly or via Talk pages, or subsequent edits made to their work.

Wikimedia, the charity behind Wikipedia, does not describe Wikipedia itself as a primary information source, but as a tertiary information source—a source for sources. By understanding the process of article creation and the issues relating to the information sources which contribute to the creation of Wikipedia articles, students will have a better understanding of both Wikipedia itself and also Wikipedia as a representative example of all information sources, all of which are subject

to the same issues of systemic bias and the subsequent consequences of symbolic annihilation.

Conclusion

It is difficult to underestimate the impact that Wikipedia continues to have on modern education and information-seeking behaviour. It is the first place many people turn to for information, both deliberately and, frequently, unknowingly. Google uses information from Wikipedia as the primary reference source for its knowledge panels (the information boxes that appear on Google when you search for people and places). Smart home assistants like Siri and Alexa draw on Wikipedia for facts and information.

The most accessed articles at any given moment can provide a good idea of what events are happening in the world at that time—for example, in the week following his death Kobe Bryant's article on Wikipedia was the most viewed article on the English language Wikipedia. Studies have even shown that Wikipedia can be used to predict outbreaks of disease, as scientists can track spikes of users in certain locations accessing information on disease-related Wikipedia pages (Generous et al., 2014).

With Wikipedia looming so large as a source of information throughout our students' lives even before they come to university, I feel it is vitally important to focus on how it functions and why. No information should be consumed uncritically, still less a source of information that has such an outsize impact on the world at large. No user can truly understand Wikipedia's strengths and weaknesses without looking 'under the hood', to see how a crowd-sourced encyclopaedia is only as neutral and impartial as the people contributing to it.

As educators, we need to train students to think critically about the information they consume, whether that information is curated for them by teachers and librarians, or available freely online on a website like Wikipedia. Students need to learn to look beyond the surface level of what an information source is (or appears to be) and what facts or opinions it contains, to the deeper history of how and why it came to be created in the first place. Wikipedia, with its transparent procedures and open-to-all approach, is an ideal tool to use to explore these issues with students.

My own experience as an educator showed me that students rarely stopped to critique or question Wikipedia as a reference source. They were

accustomed to being told not to use it in assignments or essays; when questioned on this they might say that it was because it was not a reliable source, but they could rarely expand on why that might be.

It was only when they became editors themselves and started to realise that the people creating the content they so uncritically consumed were just like them—or, as this chapter points out, quite frequently not at all like them—that they began to develop a more sophisticated understanding of Wikipedia as part of the ever-changing information cycle *process*, rather than merely as a source of information.

Information is not neutral. Libraries are not neutral. Wikipedia is not neutral. There is no such thing as a mere repository of information. At every stage in the information cycle choices are made: about what to research and write about; who and what to include and exclude; who to publish; what books to buy and stock; what is deemed worthy of study. These choices will reflect the biases, power imbalances, opinions and cultures of those involved—whether individual or institutional—conscious or not.

When my students started making those choices themselves, when they were asked to decide on a topic or person to write about in Wikipedia, they became part of that process. When we started focusing on why those people, why those topics, and discussing how our own lives and backgrounds and experiences inform the choices we make, hopefully they became part of the solution as well.

References

- Brehaut, L. (2019, October 3). 'You cannot be what you cannot see': Chef's Table star Asma Khan dishes on time-honoured Indian recipes and turning opportunity into advocacy | National Post. Retrieved 21 November 2019, from National Post website:
<https://nationalpost.com/life/food/you-cannot-be-what-you-cannot-see-chefs-table-star-asma-khan-dishes-on-time-honoured-indian-recipes-and-turning-opportunity-into-advocacy>
- Caswell, M., Migoni, A. A., Geraci, N., & Cifor, M. (2017). 'To Be Able to Imagine Otherwise': Community archives and the importance of representation. *Archives and Records*, 38(1), 5–26.
<https://doi.org/10.1080/23257962.2016.1260445>

- Centeno, A. B., Lara, J. Á. S., & Vallejo, M. D. (2014, June 24). *Analysis of the Publishing Industry in Europe*. *Enterprise Cluster*. 24. Montreal.
- Criado-Perez, C. (2019). *Invisible Women: Exposing Data Bias in a World Designed for Men* (01 edition). London: Chatto & Windus.
- Curry, M. J., & Lillis, T. (2018, March 13). The domination of English-language journal publishing is hurting scholarship in many countries (opinion) | Inside Higher Ed. Retrieved 25 November 2019, from Inside Higher Ed website:
<https://www.insidehighered.com/views/2018/03/13/domination-english-language-journal-publishing-hurting-scholarship-many-countries>
- Donaldson, K. (2019, October 22). What Big Mouth gets right and wrong about bisexuality. Retrieved 21 November 2019, from SyFy Wire website:
<https://www.syfy.com/syfywire/what-big-mouth-gets-right-and-wrong-about-bisexuality>
- Gallert, P., & van der Velden, M. (2015). The Sum of All Human Knowledge? Wikipedia and Indigenous Knowledge. In N. Bidwell & H. Winschiers-Theophilus (Eds.), *At the Intersection of Indigenous and Traditional Knowledge and Technology Design* (pp. 117–135). Santa Rose, CA: Informing Science Press.
- Generous, N., Fairchild, G., Deshpande, A., Del Valle, S.Y. & Priedhorsky, R. (2014) Global Disease Monitoring and Forecasting with Wikipedia. *PLoS Computational Biology*, 10(11): e1003892. <https://doi.org/10.1371/journal.pcbi.1003892>
- Gerbner, G., & Gross, L. (1976). Living with Television: The Violence Profile. *Journal of Communication*, 26(2), 172–199.
<https://doi.org/10.1111/j.1460-2466.1976.tb01397.x>
- Graham, M., Hale, S., & Stephens, M. (2011). *Geographies of the World's Knowledge* (C. Flick, Ed.). Retrieved from
https://www.oii.ox.ac.uk/archive/downloads/publications/convoco_geographies_en.pdf
- Klein, H., & Shiffman, K. S. (2009). Underrepresentation and Symbolic Annihilation of Socially Disenfranchised Groups (“Out Groups”) in Animated Cartoons. *Howard Journal of Communications*, 20(1), 55–72.
<https://doi.org/10.1080/10646170802665208>

- Leonard, V. (2018, December 12). Female scholars are marginalised on Wikipedia because it's written by men | Victoria Leonard. *The Guardian*. Retrieved from <https://www.theguardian.com/education/2018/dec/12/female-scholars-are-marginalised-on-wikipedia-because-its-written-by-men>
- Longest edit-a-thon | Guinness World Records. (n.d.). Retrieved 21 November 2019, from <https://guinnessworldrecords.com/world-records/430455-longest-edit-a-thon/>
- Lunch*. (n.d.). Retrieved from <https://en.wikipedia.org/w/index.php?title=Lunch&oldid=922963225>
- Talk:National Rifle Association* (n.d.) Retrieved from https://en.wikipedia.org/wiki/Talk:National_Rifle_Association
- Merskin, D. (1998). Sending up Signals: A Survey of Native American Media Use and Representation in the Mass Media. *Howard Journal of Communications*, 9(4), 333–345. <https://doi.org/10.1080/106461798246943>
- Montellaro, Z. (2015, November 18). How Does Political Wikipedia Stay Apolitical? Retrieved 22 November 2019, from The Atlantic website: <https://www.theatlantic.com/politics/archive/2015/11/how-does-political-wikipedia-stay-apolitical/450948/>
- Powell, J. (2018, October 1). Remember This? Charlotte Whitton becomes mayor [Newspaper]. Retrieved 21 November 2019, from OttawaMatters.com website: <https://www.ottawamatters.com/remember-this/remember-this-charlotte-whitton-becomes-mayor-1068110>
- Rhoden, W. C. (2018, October 26). A radical idea to solve the lack of African-American players in baseball. Retrieved 21 November 2019, from The Undefeated website: <https://theundefeated.com/features/a-radical-idea-to-solve-baseballs-diversity-conundrum/>
- Richardson, H. (2019, June 13). Women in tech are being called Tinkerbell despite their success. Retrieved 21 November 2019, from Stylist website: <https://www.stylist.co.uk/life/women-in-tech-martha-lane-fox-called-tinkerbell-despite-huge-success/272765>
- Smith, J. F. (2015, February 19). Howard University Fills in Wikipedia's Gaps in Black History—The New York Times. Retrieved 21 November 2019, from The New York Times website:

<https://www.nytimes.com/2015/02/20/us/at-howard-a-historically-black-university-filling-in-wikipedias-gaps-in-color.html>

The World Bank Group. (n.d.). Individuals using the Internet (% of population) | Data. Retrieved 21 November 2019, from World Bank website:

https://data.worldbank.org/indicator/IT.NET.USER.ZS?most_recent_value_desc=false

Tuchman, G. (1978). The symbolic annihilation of women by the mass media. In A. Daniels & J. Benet (Eds.), *Hearth and home: Images of women in the mass media* (pp. 3–38). New York: Oxford University Press.

Venzo, P. (2016, September 1). To boldly go where we should have gone before: Symbolic annihilation and queer interventions in the Star Trek textual universe. https://doi.org/10.1386/qsmc.1.3.285_1

Wikipedia:How to run an edit-a-thon. (2019). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Wikipedia:How_to_run_an_edit-a-thon&oldid=915344486

WikiProject: Countering systemic bias. (n.d.). Retrieved from https://en.wikipedia.org/w/index.php?title=Wikipedia:WikiProject_Countering_systemic_bias/Geography&oldid=905329260

(N.d.). Retrieved 25 November 2019, from Ethnologue website: <https://www.ethnologue.com/language/eng>